

A Study on Big Data

A. Subha¹, S. Joshna², P. Banu Sree³

¹Student, Department of Computer Technology, Sri Krishna Arts and Science college, Coimbatore, India

^{2,3}Asst Professor, Dept. of Computer Technology, Sri Krishna Arts and Science college, Coimbatore, India

Abstract—BIG DATA is a process of extracting a data from large dataset. Challenges of big data includes capturing information, information storage, information analysis, search, sharing, updating, visualizing, data privacy, information supply. Challenging task in analyzing a big data involves massive distributed file systems that ought to be fault tolerant, versatile and ascendable. The size of information sets being collected and analyzed within the business for business intelligence is growing apace, creating ancient storage solutions prohibitively costly. Hadoop could be a fashionable ASCII text file map-reduce implementation that is being employed in corporations like Yahoo, Facebook etc. to store and method very giant information sets on artefact hardware. This paper aims to analyze some of the different characteristics, applications of big data is presented.

Index Terms—dataset, storage, 5v's, devices, privacy, massive data.

I. INTRODUCTION

As the, large data volumes are generated daily from health, government, social media, financial, marketing. This is because of technological trends including IoT as well. Big data is nothing but data with a huge large in size of data. 'Big data' is a term which is used to describe a collection of data (i.e.) large in size. An accurate definition of 'big data' is difficult to nail down because projects, vendors, practitioners and business professionally use it quite differently. In generally big data is nothing but large dataset and to handle large dataset it use some category of compulsory strategies and technologies.

Use some category of computing strategies and technologies spread of smart devices. Big data is a new concept for handling the massive data. Big data analysis the large amount of information which is very useful and important information for the use. The need of big data analysis that is the processing on data sets which are more complex and massive. Such data is different from the structured data with five parameter, the parameter are variety, volume, values veracity and velocity. These five parameter are simply known as (5v's) five v's, these are the characteristics of big data [6].

II. CHARACTERISTICS OF BIG DATA

A. Velocity

Velocity is refer to the speed of vast amounts of data which is being generated, collected and analysis. In daily days the number of messages, emails, photos, video's etc... are increasing day by data. We can analysis the data while it is

generated by using big data technologies [6].

B. Volume

Volume is refers to the amount of data which is generated by social media, cell phone, cars, credit cards , video, sensors, etc. The data is collected and analysis is clearly an engineering challenge of immensely vast proportions [6].

C. Value

If we talk about value, we are referring to the worth of the data being extracted. Having endless amount of data is one thing, but unless it can be turned into values it is useless [6].

D. Variety

The variety is defined as the different type of data. The new technologies of big data allow both Structured and unstructured data to be stored and used simultaneously [6].

E. Veracity

Veracity is the quality of the data. The results of big data may be typically directly measured mating it simply to determine a return investment. The massive knowledge may be a tool undoubtedly price wanting into [6].

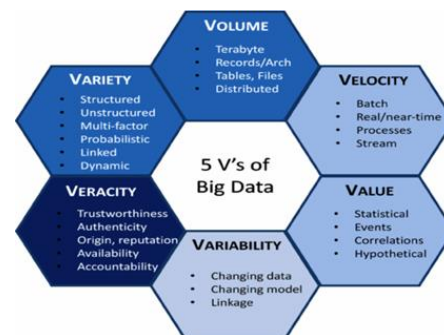


Fig. 1. Characteristics of big data

III. APPLICATIONS OF BIG DATA

A. Communications, Media and Entertainment

Many organizations, industries need big data Support. In suburban area Championship (You Tube Video) big Data can take full advantage to bring analysis on the lawn tennis, cricket, reality shows, and films in TV, and Mobile in real time. A music service is served on-demand, wont to collect all relevant knowledge from its billions of users in real time worldwide. This knowledge are analyzed and wise music recommendations

to individual users [1], [2].

B. Financial Industry

All the monetary trade can pay the cash to require to use of data, to form higher profit Hadoop is getting used in the trade for sentiment analysis, prophetic analysis, and the monetary trades [1].

C. Supply Chain, Logistics and Industrial Engineering

Companies are using telematics and big data to contour transport fleets and the way they will improve fuel usage and routes. Union Pacific Railroad use thermometers, microphones, and ultrasound to capture information concerning their engines and send it for analysis to spot instrumentality in danger for failure [1].

D. Education

By victimization big data, students have deployed a learning and management systems. Manipulating this situation once the student or school logs on to the their portal or system, it'll show the small print concerning their operating hours and pages [1], [2].



Fig. 2. Applications of big data

IV. BENEFITS OF BIG DATA

A. Cost Savings

Some tools of big data like Hadoop and Cloud-Based Analytics will bring value benefits to business once massive amounts of information area unit to be keep and these tools additionally facilitate in characteristic additional economical ways that of doing business[3], [4].

B. Time Reductions

The high speed of tools like Hadoop and in-memory analytics will simply establish new sources information that helps businesses analyzing data straight off and build fast choices supported the learnings[3], [4].

C. Fraud Detection

Fraud is detected the instant it happens and correct measures is taken to limit the injury. The financial world is extremely enticing for criminals. With a time period safeguard system, tries to hack into your organisation area unit notified instantly [3], [4].

D. Control Online Reputation

Big data tools will do sentiment analysis. Therefore, you'll get feedback concerning UN agency is oral communication what concerning your company. If you wish to observe and improve the web presence of your business, then, big data tools will facilitate this [3], [4].

V. CONS OF BIG DATA

A. Storage Issues

Storage is extremely complicated," Robinson says. And indeed, not solely will it entail managing capability and determining the most effective assortment and retrieval ways, it additionally means that synching with each the IT and also the business groups and listening to complicated security and privacy problems [7].

B. Management Issues

The most troublesome downside in big data is management is economical organization, administration and governance of information sets. we'd like to follow protocols for making certain knowledge accuracy and validity. Finding outliers could be a major task of the information qualification [7].

C. Processing Issues

For process big data, timeliness is a vital purpose to be thought of. We tend to need data processing and new analytic algorithms for ensuing unjust data. Data processing techniques that currently directly apply for intra-node similarly, which suggests hardware resources appreciate processor caches and processors memory channels area unit shared across one node. There are a unit several things wherever the results of the analysis is needed forthwith [7].



Fig. 3. Pros and Cons of big data

VI. CHALLENGES OF BIG DATA

Big data investigation desires different types of stages that incorporate info acquaint, information extraction and cleanup, info combination, conglomeration and illustration, inquiry process, info demonstrating and examination, understanding. Each of above had challenges pertinent with non-uniformity, scale, timeliness, multifarious nature and protection [8].

A. Heterogeneity

Big information regarding the large and an oversized quantity of resources; this suggests a distinct sort of information

captured and picked up for analysis method, the algorithms that used for big data analysis method forever handling same information and can't perceive non-uniformity. Let's say, the audio and video information with traditional information. This ends up in type and construct the information before data analysis fastidiously [1].

B. Size of the Data

The name itself "Big Data", the degree of the info represent a significant challenge since the info is growing apace. The method of managing massive size of information needs an efficient resolution to handle this challenge [1].

C. Processing Time

Big data processing takes longer than traditional processing. The sure-fire system suggests that to supply the info to the user properly and on time. With a big data, it's tough to handle an oversized quantity of knowledge in term of interval [1].

D. Sharing of Data

Information sharing would support the provision of the data publically, however users refuse to share their big data for competitive purpose. Suppose that every user share their own big data a lot of correct result is often obtained so as style new and customary standardization, to handle every information imported from totally different sources and platforms [1].

E. Privacy

One of the essential challenges that has got to be rigorously managed is privacy. Big data stores differing kinds of data as well as public, private, and private data. additionally to nobody wish to share their personal data like secret, mastercard variety, health standing report....etc. to avoid misuse of this valuable information. For this reason, big data should realize associate economical resolution to guard user privacy, particularly for private data it ought to give multi lines of defense relating to any expecting attack or misuse of user personal data [1].

VII. CONCLUSION

Big data typically includes information sets with sizes on the far side the power of unremarkably used code tools to capture, curate, manage, and method information at intervals a tolerable time period. Big data philosophy encompasses unstructured, semi-structured and structured information, but the most focus is on unstructured information. Massive information "size" could be a perpetually moving target, as of 2012 starting from many dozen terabytes to several Exabyte's of information. Big data needs a collection of techniques and technologies with new varieties of integration to reveal insights from datasets that area unit various, complex, and of a vast scale. In this paper we had discussed about applications, characteristics, challenges, pros and cons of big data. In future this paper is enhanced by the topic of big data analysis.

REFERENCES

- [1] R.Gokilabharathi, K.Gurulakshmi," Prominence of Big data and its related issues, challenges and applications", International Journal on Computer Science and Engineering (IJCSSE).
- [2] <https://intellipaat.com/tutorial/hadoop-tutorial/big-data-overview/>
- [3] <https://www.newgenapps.com/blog/importance-benefits-competitive-advantage-big-data>
- [4] Bala M. Balachandran and Shivika Prasad,"Challenges and Benefits of Deploying Big Data Analytics in the Cloud for Business Intelligence", International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France.
- [5] P.Sai Prasad, K.Rajesh," A Novel Study on Big Data: Issues, Challenges, Tools", IJCSET, 2017 | Volume 7 Issue 6, 44-49.
- [6] C. Lakshmi , V. V. Nagendra Kumar,"Survey Paper on Big Data". Volume 6, Issue 8, August 2016, International Journal of Advanced Research in Computer Science and Software Engineering.
- [7] Rachana Sharma , Priyanka Sharma," A Novel Approach towards Big Data Challenges", International Journal of Innovative Computer Science & Engineering Volume 1 Issue 2; Page No. 28-34.
- [8] Raswitha Bandi, Dr. Sheikh Gouse, Dr. J.Amudhvel," A Comparative Analysis for Big Data Challenges and Big Data Issues Using Information Security Encryption Techniques, International Journal of Pure and Applied Mathematics Volume 115, No. 8, 2017, 245-251.



Fig. 4. Challenges of big data